ICSG 2011





Structural and functional genomics of a model organism *Thermus thermophilus* HB8: toward functional discovery of functionally unknown proteins (Poster 135)

Akeo Shinkai Team Leader, SR system Biology Research Group (Group director: Dr. Seiki Kuramitsu)

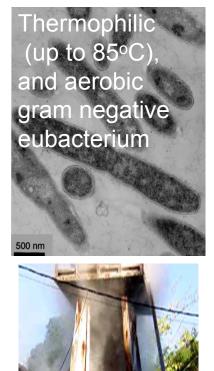
RIKEN SPring-8 Center, JAPAN

1. Whole Cell Project of *T. thermophilus* HB8

- 2. Structural Genomics
- 3. Functional Genomics
- 4. Resource and Database

Whole cell project of *T. thermophilus* HB8

The ultimate goal of this project is to understand all of the fundamental biological phenomena at an atomic resolution, firstly, focusing on proteins.



The reasons why *T. thermophilus* are (1)~2.1 Mb genome and ~2,200 genes (half of *E. coli or Bacillus subtilis*)

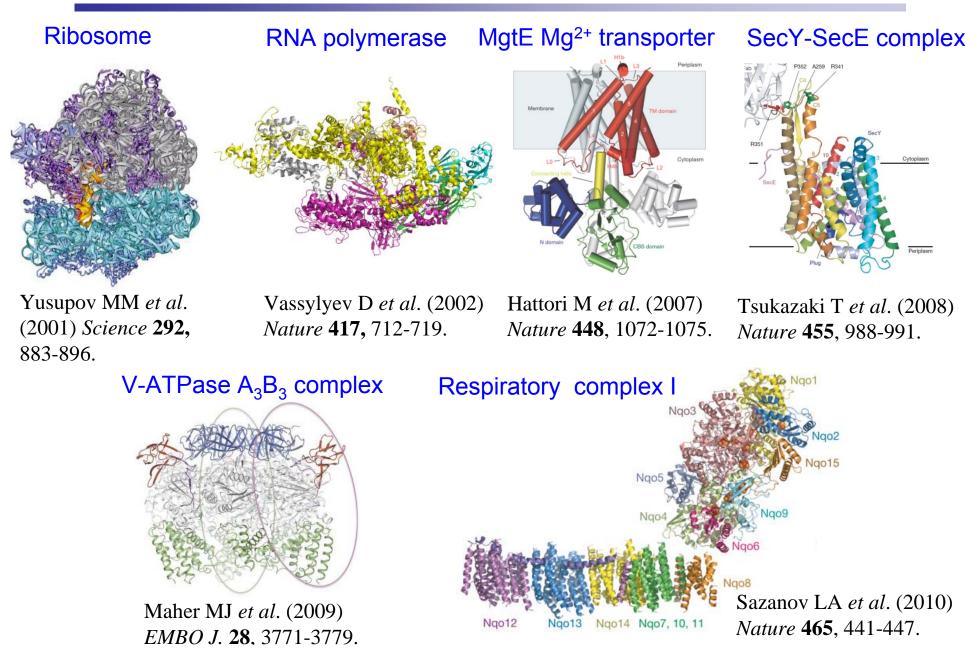
(2) It can grow in a minimum medium.

(3) Basic genetic engineering techniques on this strain are established. (construction of gene disruptant strain, expression of recombinant protein)

(4) Many proteins from this strain are heat stable.(Suitable for their structural and functional analyses)

Isolated by Dr. Tairo Oshima from "Mine" hot springs in Japan

Crystal structures of large complexes and membrane proteins



4/26

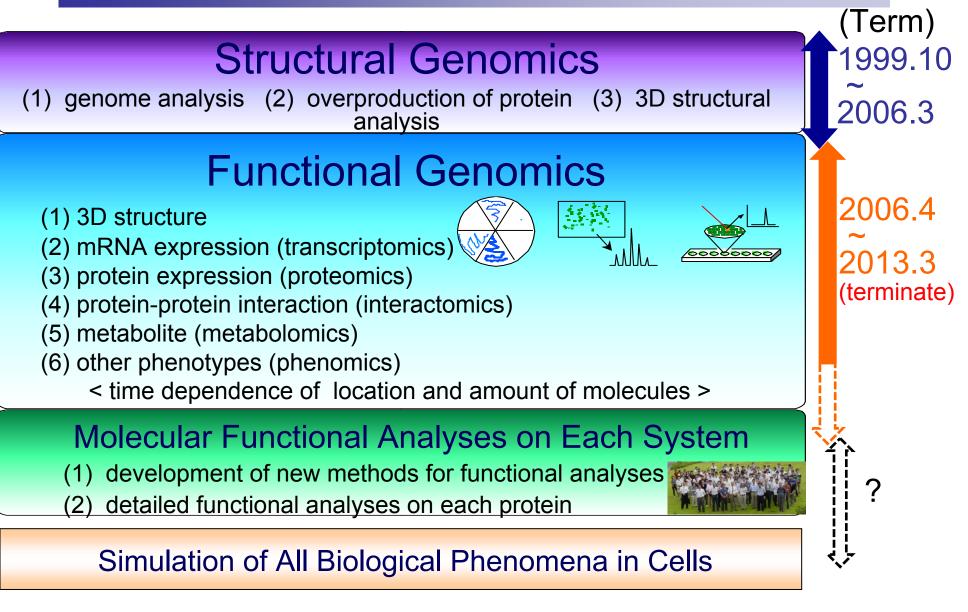
ISGO	International Stru UserID (E-mail)	octural Genomics Organizati	Genom	e decoding project ama T & Shibata T)
HOME	Brief histor	y of the international structura	genomics projects	
<u>History</u>		,	8 F)	
<u>Organization</u>	1995	Proposal of structural genomics pro	oject in Japan	
<u>ISGO Reports</u> and Agreements	1997 Apr.	Start of structural genomics pilot p	roject at RIKEN Institute	Whole cell project
Project List	1997	Initiating study of structural genomi	cs at DOE and NIGMS/N	(Kuramitsu S)
<u>Meetings /</u> Conferences	1997	<u>Start of the New Jersey Initiative in</u> <u>Bioinformatics</u>	Abstracts of the 7th Annual Meeting Tokyo, May 1-2, 1995	of the Protein Engineering Society of Japan
Workshop	1997 Jan.	<u>The workshop on Structural Genom</u>	40. Database of heat-stable proteins from <i>Thermus thermophilus</i> HB8	weight and isoelectric point of each protein spot were calculated. The average molecular weight and isoelectric point values were estimated to be 30 000 and 5.2, respectively. After the protein spots had been electroblotted onto a polyvinylidene
<u>Other topic-</u> related mtgs	1998	Start of the initial pilot projects in (Sciki Kuramitsa, Shin-ichi Kawaguchi and Yoshiaki Hiram Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan	
Membership	1998 Oct.	Structure-Based Functional Genom	Thermus thermophilus HB8 is a Gram-negative eubacter which can grow at temperatures of up to 85°C. It is the r thermophilic bacterium known to have a system of ger	nost sequence. We are also constructing a genome DNA bank of
<whole cell="" p<="" th=""><th>oroject in RIK</th><th>Coll for grant applications for NICM</th><th>engineering. Proteins from this organism are known to thermostable and have several advantages for study as folk (i) They are stable for a long period at room temperature are resistant to proteases. (ii) If cloned and expressed mesophilic bacterium, heat treatment would be a very effec method for their purification. (iii) They have a good proper to form crystals. Thermostable proteins from <i>Thermoph</i></th><th>be probably one of the smallest genomes possessed by an autotroph. This small genome would thus be useful for analyzing the structure-function relationships of all the molecules in a single specified autotrophic cell. Therefore, <u><i>Thermophilus</i></u> would be a better candidate for studying the two distructure functions living cell at the molecular biology.</th></whole>	oroject in RIK	Coll for grant applications for NICM	engineering. Proteins from this organism are known to thermostable and have several advantages for study as folk (i) They are stable for a long period at room temperature are resistant to proteases. (ii) If cloned and expressed mesophilic bacterium, heat treatment would be a very effec method for their purification. (iii) They have a good proper to form crystals. Thermostable proteins from <i>Thermoph</i>	be probably one of the smallest genomes possessed by an autotroph. This small genome would thus be useful for analyzing the structure-function relationships of all the molecules in a single specified autotrophic cell. Therefore, <u><i>Thermophilus</i></u> would be a better candidate for studying the two distructure functions living cell at the molecular biology.

1. Structurome Research Group, FY1999 ~ 2006 the gel with silver standard marker proteins, the molecular Using comigrated standard marker proteins, the molecular

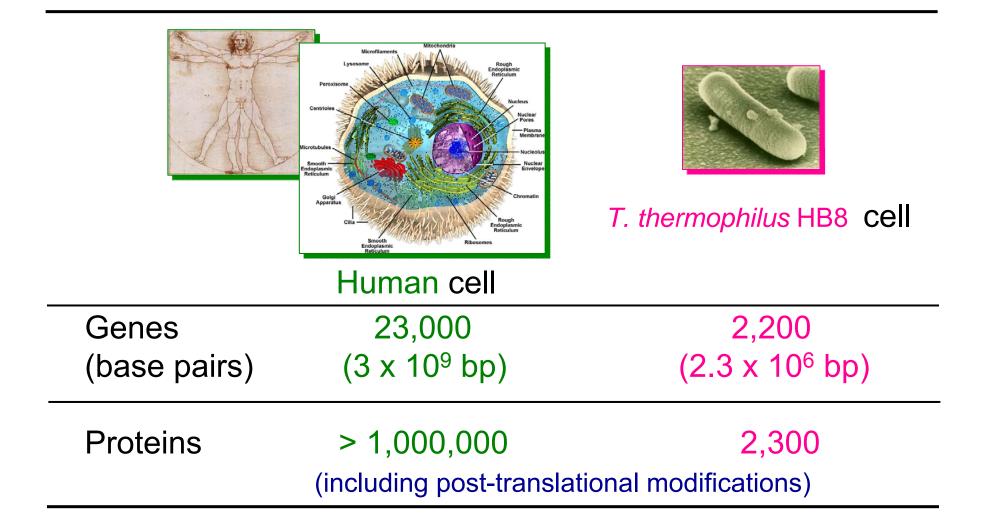
(Group director: Kuramitsu S & Yokoyama S) (\$ 2 million/year

- Kuramitsu, S. et al. (1995) Protein Eng. 8, 964
- 2. RIKEN Structural Genomics/Proteomics Initiative, 2001 National Project on Protein Structural and Functional Analyses, "Protein 3000", FY2002 ~ 2006
- 3. SR System Biology Research Group, FY2006 ~ 2012 (terminate) (Group director: Kuramitsu S.) (now \$ 1 million/year)

Long-term strategy of the Whole-cell project in RIKEN



Structural and functional genomics of *T. thermophilus* HB8



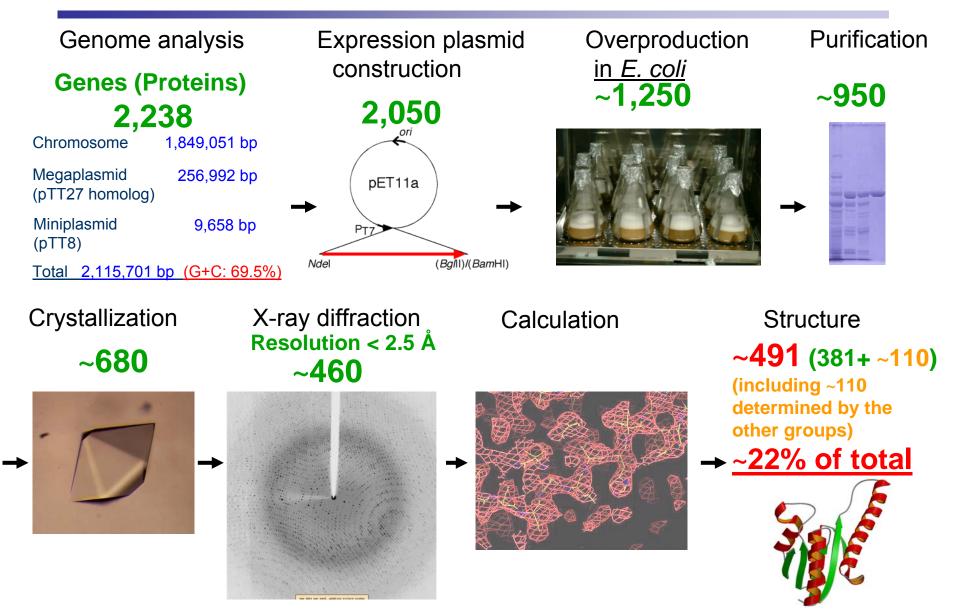
With this model organism, we hope that basic biological phenomena common to many organisms, including human will be elucidated.

1. Whole Cell Project of *T. thermophilus* HB8

2. Structural Genomics

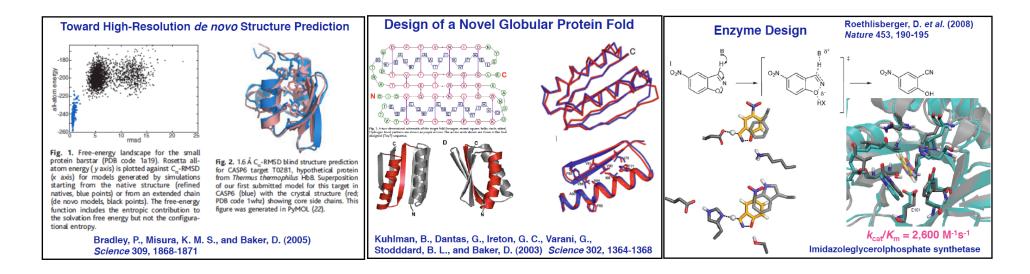
- 3. Functional Genomics
- 4. Resource and Database

Structure determination of the proteins



This strain is one of the organisms whose structural genomics are much progressed.

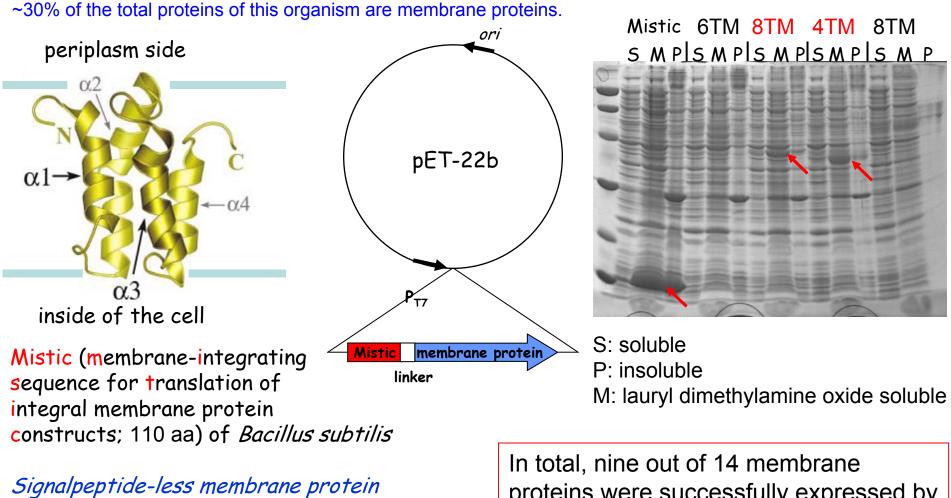
Prediction and *de novo* design of protein structures



Internationally cooperative efforts in protein structure determination increased the success rate of the protein backbone conformations to about 70%.

The *T. thermophilus* protein structures also contribute to the development of programs for prediction or *de novo* design of protein structures.

Trial expression of membrane protein



[Roosild, T. P. *et al.* (2005) *Science* **307**, 1317-1321]

proteins were successfully expressed by this system.

This expression system might be useful to obtain large amounts of various membrane proteins with high efficiency.

Topic

1. Whole Cell Project of *T. thermophilus* HB8

2. Structural Genomics

3. Functional Genomics ~ toward functional discovery of functionally-unknown proteins
4. Resource and Database

"30~40% of total proteins are hypothetical (functionally-unknown) proteins."

T. *thermophilus* has many functionally unknown proteins

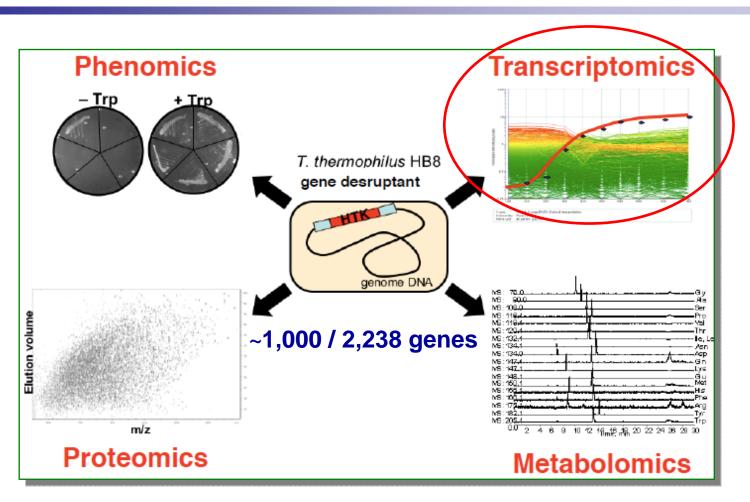
COG code	Description	No. in genome	
	Poorly characterized		
R	General function prediction only	304	
S	Function unknown	166	
-	Not in COGs	434	E
			ι

According to the **Clusters of Orthologous Group of proteins (COG)-based categorization**, 600 **functionally-unknown proteins (genes)** are found in this strain.

Elucidation of function of the functionallyunknown proteins is necessary for an understanding of the whole cell life system.

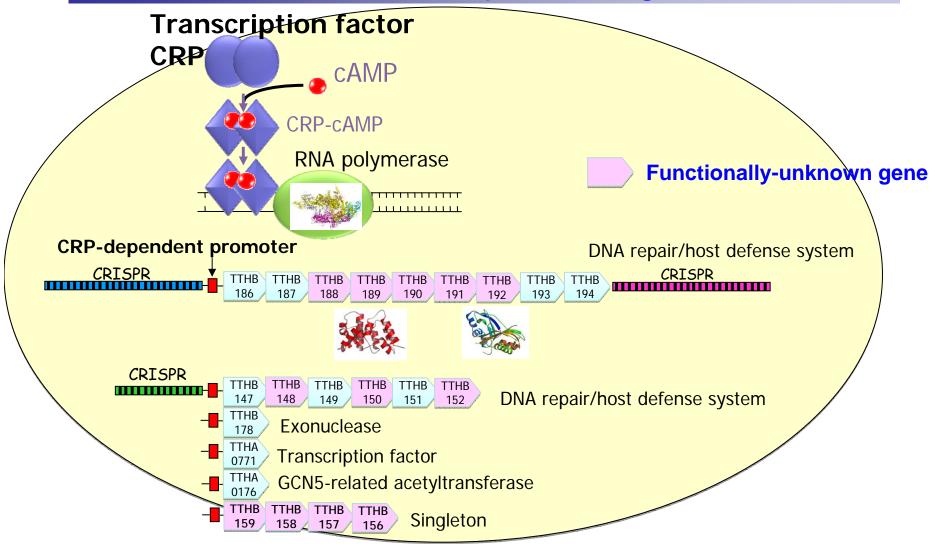
Strain	COG code S	Not in COGs	Total
T. Thermophilus HB8	166	434	600
<i>E. coli K-12</i> (W3110)	322	585	907
B. subtilis (str.168)	340	900	1,240

Construction of the platforms for functomics analysis



Classify the functionally-unknown proteins (genes) based on their transcriptional regulation and obtain clues as to their function.

Classification of the functionally unknown gene (protein) based on transcriptional regulation



Transcription of several genes sharing similar cellular function is often synchronously regulated.

Study of transcription using *T. thermophilus* HB8

Strain	Genome (Mbp)	Number of gene	Number of transcription factor	Number of σ factor
T. thermophilus HB8	2.1	2,200	~70	2
Escherichia coli	4.7	4,300	350	7
Bacillus subtilis	4.3	4,100	330	17

T. thermophilus HB8 is an appropriate model organism to study fundamental transcriptional regulatory system.

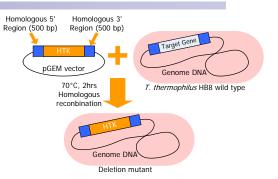
Strategy for functional identification of transcription factor (TF)

[A] Molecular function

- a) Identify target genes of TF
 - **DNA microarray (transcriptome) analysis** Compare total mRNA expression of TF gene-disrupted strain with that of wild type.
 - · Genomic Selex
 - · In vitro transcription analysis (and promoter search)
- b) Determine three-dimensional structure
 - · X-ray crystal structure









17/26

FadR-lauroyl-CoA

[B] Cellular function (Physiological function)

- a) Analyze altered mRNA expression caused by environmental alteration
 - DNA microarray analysis
- b) Analyze function of the target gene products (proteins)
 - Activity measurement, prediction from amino acid sequence or X-ray crystal structure

Summary of the number of the target gene of <u>*T. thermophilus* regulators</u>

Regulator	No. of the target promoter	No. of the target gene	Regulator	No. of the target promoter	No. of the target gene
NusG	(for the activity	of RNAP)	CRP	6	22 (12)
σ ^A	(housekeepin	g genes)	SdrP	16	22 (6)
GreA	(for the activity	of RNAP)	FadR	9	21 (2)
MIC	1	3	PaaR	2	11 (3)
NusA	(for the activity	of RNAP)	CsoR	1	3 (1)
Gfh1	(for the activity	of RNAP)	σ ^E /anti-σ ^E	3	5 (4)
ArgR	2	5 (1)	TTHB099/ LitR	2	5 (0)
SIrA	1	1	Total	43	98 (29)
SIpM	1				

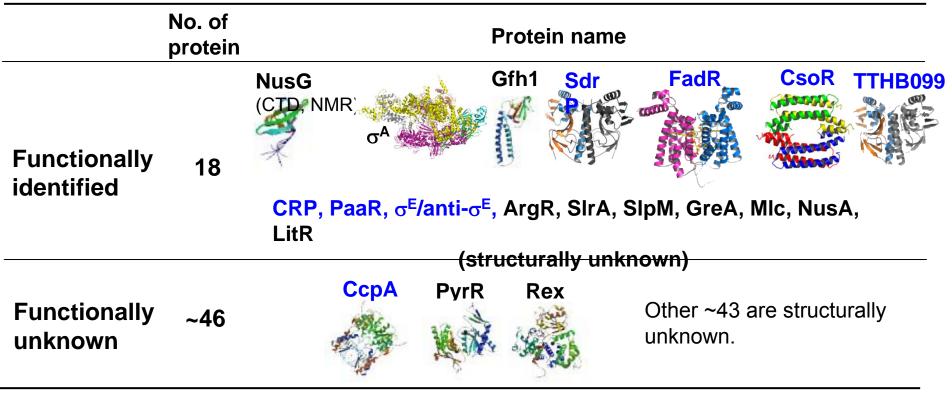
: studied by our team

18/26

(): number of functionally unknown (COG code S or non-categorized) gene

So far, in total, 98 genes containing 29 functionally-unknown or hypothetical genes out of \sim 2,200 could be categorized based on the activity of them.

T. thermophilus transcriptional regulator



(Blue letter: studied by our team)

Functional identification of remaining all transcription factors is necessary to classify remaining functionally-unknown proteins.

- 1. Whole Cell Project of *T. thermophilus* HB8
- 2. Structural Genomics
- 3. Functional Genomics
- 4. Resource and Database

Plasmids

RIKEN BIORESOURCE CENTER

http://www.brc.riken.jp/lab/dna/en/thermus_en.html

B I	RRIKEN ORESOURCE CENTER	R I K E	NBI	o r e s o				200 MA	Mulul		
		Hor	ne	Cle	one set	Gei	ne Expression	Gene Analys	sis Se	arch	Literature
E.	coli Sc. pombe	S. cerevisia	e T. th	nermophilus	Other M	icrobe	Mammalian Co	ells			
E	xpression in Micro	obe									
Т	hermophili	ic Micr	oorg	anisms	s Gene	Plas	smid				
		pernix and S.	0								
I	hermus the	rmoph	ilus g	gene pl	asmid						
Or	tline (ja) Outline	(en) Expr	ession P	lasmid D	isruption Pl	asmid	Table of summa	uy			
The	utline rmus thermoph rmus thermoph		-							-	
	earch Group, R					•		•			- 0,
		V.		1	Г		<u> </u>				
	~2,050	clon	es			~1	,000 c	lones			

DNA microarray data

NCBI Gene Expression Omnibus (GEO)

(http://www.ncbi.nlm.nih.gov/projects/geo/)

S NCBI	Gene Expression Omnibus	
HOME SEARCH SITE MAP	Handout NAR 2006 Paper NAR 2002 Paper	FAQ MIAME Email GEO
	Omnibus: a gene expression/molecular abundance ng MIAME compliant data submissions, and a curated,	Not logged in Login 2
	ene expression data browsing, query and retrieval.	GPL Platforms 4951 GSM Samples 248905 GSE Series 9551 Total 263407
QUERY	Gene profiles GO	Site contents Documentation
Browse	GEO accession GO GEO BLAST DataSets Platforms GEO accessions Samples	Overview FAQ Submission guide Linking & citing Journal citations Programmatic access DataSet clusters GEO announce list Data disclaimer GEO staff
	Series	Ouery & Browse 12 Repository browser Submitter contacts SAGEmap FTP site GEO Profiles
SUBMIT	Direct deposit / update Create new account Web deposit / update	GEO DataSets Deposit & Update @ Direct deposit Web deposit
	een elements for information	New account
Get GEO accession Depositors only	User : Password : COCIN Recover	unt: Quick 💌 🔽 🛛 🗤 Quick 💌

Platform: accession number GPL9209

418 samples,58 experimental series

Open access homepage of the whole cell project

http://www.thermus.org/

Control Andrea Control Andrea Planet Contractor Control Antonia Control Antonia Control Antonia Control Antonia	Structural-Biologic Whole Cell Projec		
32 Shatan Anave antainat	Jug 23, 2010 Thank you very such for aching to the Sch Uncting of De	ole Cell Project of TMSS. The group platax and	
100			
althrough 2	!Attention! If you are using Google Chrome or a fast brow: "Reload" on the pages for viewing the current p		
	PThe 9th Annual Meeting of Whole Ce	II Project of TTHB8	
	Generatives, S. et al., (2011) O visualization and preliminary X-ray 3 physiology version physionatases it and Thermal Decompilation MB Societ F Societ Block Cryst Commune 57 + 390-5 Graduer, B. et al., (2011) Normic tessel Description of the commune	27 Ada Crystallogr	
	Oprotein expression vectors are now available from BRC	link to BIOR	SOURCE CENTER
	INAL SAME	IKEN Bloresource Center (BRC).	
	 http://www.birc.itken/Jpi/do/Afnatien/Thermus_en/Jhml All clones are consituated in p ET expression system without Clone set: Clones are shipped as tooen recombinant 8, coli Clone (clone: Clone(3) is shipped as DNA solution in TE bur 	i lag residues. D HSa in S6well plates with dry-ice .	

Whole-Cell Project Database

Japanese		WCPDB - Thermus thermophilus HB8 Whole Cell Project Database -
		Database Target Search Mode Article Search Mode
		Genome Map chromosome (TTHA) mega plasmid (TTHB) mini plasmid (TTHC)
		Pathway Pathway with Structure
\frown	т	ranscriptome datasets of <i>T. thermophilus</i> HB8 GeneChip Custom Array are available on the NCBI GEO.
Accession	Subio File	Description
GPL9209	GPL9209.tab	GeneChip TTHB8401a520105F Platform Rev.2 < The new library for custom array was updated on September 18 2009 !>
GPL4902	GPL4902.tab	GeneChip TTHB8401a520105F Platform Rev.1
GSE7165		Time course of the mRNA expression in wild-type Thermus thermophilus HB8 strain grown on a rich medium (labeling:ENZO)
GSE7166		Comparative expression analysis between crp (TTHA1437) deletion mutant and wild-type of T. thermophilus HB8
GSE7175		SuperSeries for the study of expression analysis of the T. thermophilus CRP (TTHA1437)
GSE8780		Comparative expression analysis between sigE (TTHB211) deletion mutant and wild-type of T. thermophilus HB8
GSE8781		SuperSeries for the study of expression analysis of the T. thermophilus sigE (TTHB211)
GSE8795		mRNA expression in TTHA1939 deletion mutant of Thermus thermophilus HB8 strain grown on minimum medium
GSE10368	GSI 10368.ssa	Time course of the mRNA expression in wild-type Thermus thermophilus HB8 strain grown on a rich medium
GSE10369		Comparative expression analysis between sdrP (TTHA1359) deletion mutant and wild-type of T. thermophilus HB8
GSE10370		SuperSeries for the study of expression analysis of the SdrP (TTHA1359) in T. thermophilus HB8
GSE10371		Comparative expression analysis between TTHB212 deletion mutant and wild-type of T. thermophilus HB8
GSE10548		SuperSeries for the study of expression analysis of the T. thermophilus sigE (TTHB211) and anti-sigE (TTHB212)
GSE11671	SSE11671.ssa	Comparative expression analysis between logarithmic and stationary phase of <i>T. thermophilus</i> HB8 on a synthetic medium
GSE12590		mRNA expression in TTHA0118 deletion mutant of <i>Thermus thermophilus</i> HB8 strain grown on minimum essential medium
	Subio Platfo	rm

FREE software for omics data management and visualization

Free Download

Whole-Cell Project Database

Search Target	Whole Cell Proje	ect Database -Targ	get List-					
Target Article	Show All List	liew						
Species: 🔷 🍸 T. thermophilus 💙	First Prev Record	s1 - 50 / 8243 Next	Last					
Locus V								
product: inc. V	Click Column Labe	el for Change Sort Or	rder				DNA	Structure
PDB == 💌	.ocus Synonyn Tag	n Product	Locus	Category	Clone for (Expression D	clone for M isruption	icroarray Data	Status PDB Deposit
O "OR" search ③ "AND" search	TTHA0001 TT0744	DNA polymerase III, beta subunit	1:551182	replication	available	available		
Display: 50 💌 Records Search Reset	THA0003 TT0470	pyruvate kinase	1:24903914	energy metabolism	available			
	TTHA0004 TT2019	hypothetical protein	1:complement(39154895)	hypothetical proteins	available			
<i>T. thermophilus</i> 188 Genome Map chromosome pTT27 pTT8	TTHA0005 TT1522	metallo- beta-lactamase family protein	1:complement(48925506)	conserved hypothetical (ZZ)	available	available		
	TTHA0006 TT0233	1-deoxy-D-xylulose- 5-phosphate synthase	1:complement(54937340)	lipid metabolism	available			
	TTHA0007 TT1523	hypothetical protein	1:complement(73497993)	conserved hypothetical (ZZ)	available	available		
	TTHA0008 TT1141	phage shock protein A	1:complement(80378699)	annotated others	available	available		
	TTHA0010 TT1993	hypothetical protein	1:90229114	hypothetical proteins				
Soarch vour	TTHA0011 TT0257	molybdenum cofactor biosynthesis protein A (MoaA)	1:1249313503	coenzyme biosyntheses	available			
Search your	ттнаоо12тт1913	conserved hypothetical protein, integral membrane protein	1:complement(1348114311)	conserved hypothetical (ZZ)	available	available		
	TTHA0013 TT0153	geranylgeranyl diphosphate synthetase (GGPS)	1:1432915321	annotated others	available			1WMW 07/21/2004
	TTHA0014 TT2225	hypothetical protein	1:1550816104	hypothetical proteins				
	TTHA0015 TT0238	LytB protein	1:1610617128	lipid metabolism	available			
	TTHA0016 TT1912	tRNA-dihydrouridine synthase	1:1712118149	modification enzymes	available	available		
http://www.brc.riken.jp/lab/dna/en/thermu:	TTHA0017 TT1911 s 3en.html	TPR repeat- containing protein	1:complement(1804518677)	conserved hypothetical (ZZ)	available	available		

Whole Cell Project Database

Search Target	Whole Cell Project Data	abase -Purification-						
Target Article	Show All List View							
Species: 🔨 🗘 T. thermophi	Thermus thermophilus HB8	TTHA0001						
Locus 🔨 🛊	DNA polymerase III, beta subunit							
Tag:	Lot Information:	Final SDS	-PAGE:					
product: inc. 🛊	Purification Lot: PC010001-4	1-01-01	1					
PDB: == \$	Plasmid ID: PC010001-41							
F 00 V	Plasmid Vector: pET-11a							
🔘 "OR" search 💽 "AND" search	Label: Native	D) M						
	Person: Inoue Y.	9	7					
Display: 50 🗘 Records	Purified Date: 07/17/2003		-					
Search Reset	Total Protein Yield: 42.3 mg	6						
	Final Volume: 2.1 mL	4	5					
T. thermophilus HB8 Genome	Conc.: 20.1 mg/mL		(Care	÷-				
Мар	Culture Lot:							
chromosome pTT27 pTT8	date host cell scale							
	05/09/2003 BL21(DE3) 6 L	1	5					
	Cell Lysis and Heat Treatme	nt:						
	18.5g of E. coli cells							
	Ultra Sonication with 20mM Tr	ris-HCl + 50mM NaCl (pH8.0)						
	1							
	Heat Treatment at 70°C for 10 min							
	↓							
	Ultra Cengrifuge at 200,000G	for 1 hr						
	Purification Procedure:							
	Column	Buffer	Flow Rate (mL/min)	Eluted				
	Resource ISO 6ml	50 mM NaPi (pH 7.0), 1.2 M (NH4)2SO4	5	0.56 M AS				
	Resource Q 6ml	20 mM Tris-HCl (pH 8.0)	5	0.2 M NaCl				
	Hydroxyapatite CHT10-I	10 mM NaPi (pH 7.0)	4	50 mM NaPi				
	HiLoad 16/60 Superdex 75 pg	20 mM Tris-HCl (pH 8.0), 150 mM NaCl	0.5					

HiPrep 26/10 Desalting 53ml 20 mM Tris-HCl (pH 8.0)

7